Data Analyst Project-Hotel Booking

October 3, 2023

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

[2]:
```
df= pd.read_csv("D:\\DataScience\\Projects\\Data Analysis\\Hotel␣
 ↪Booking\\hotel_booking.csv")
df
```

[2]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number
0	Resort Hotel	0	342	2015	July	27
1	Resort Hotel	0	737	2015	July	27
2	Resort Hotel	0	7	2015	July	27
3	Resort Hotel	0	13	2015	July	27
4	Resort Hotel	0	14	2015	July	27
…	…	…	…	…	…	…
119385	City Hotel	0	23	2017	August	35
119386	City Hotel	0	102	2017	August	35
119387	City Hotel	0	34	2017	August	35
119388	City Hotel	0	109	2017	August	35
119389	City Hotel	0	205	2017	August	35

```
        arrival_date_day_of_month  stays_in_weekend_nights  \
0                               1                        0
1                               1                        0
2                               1                        0
3                               1                        0
4                               1                        0
…                             …                        …
119385                         30                        2
119386                         31                        2
119387                         31                        2
119388                         31                        2
119389                         29                        2

        stays_in_week_nights  adults  …  customer_type     adr  \
0                          0       2  …      Transient    0.00
1                          0       2  …      Transient    0.00
2                          1       1  …      Transient   75.00
3                          1       1  …      Transient   75.00
4                          2       2  …      Transient   98.00
…                        …     …  …              …     …
119385                     5       2  …      Transient   96.14
119386                     5       3  …      Transient  225.43
119387                     5       2  …      Transient  157.71
119388                     5       2  …      Transient  104.40
119389                     7       2  …      Transient  151.20

       required_car_parking_spaces total_of_special_requests  \
0                                0                         0
1                                0                         0
2                                0                         0
3                                0                         0
4                                0                         1
…                              …                         …
119385                           0                         0
119386                           0                         2
119387                           0                         4
119388                           0                         0
119389                           0                         2

       reservation_status reservation_status_date               name  \
0               Check-Out              2015-07-01       Ernest Barnes
1               Check-Out              2015-07-01        Andrea Baker
2               Check-Out              2015-07-02      Rebecca Parker
3               Check-Out              2015-07-02        Laura Murray
4               Check-Out              2015-07-03         Linda Hines
…                     …                       …                  …
119385          Check-Out              2017-09-06     Claudia Johnson
```

```
119386          Check-Out              2017-09-07      Wesley Aguilar
119387          Check-Out              2017-09-07        Mary Morales
119388          Check-Out              2017-09-07  Caroline Conley MD
119389          Check-Out              2017-09-07      Ariana Michael

                             email  phone-number       credit_card
0       Ernest.Barnes31@outlook.com  669-792-1661  ************4322
1            Andrea_Baker94@aol.com  858-637-6955  ************9157
2        Rebecca_Parker@comcast.net  652-885-2745  ************3734
3                 Laura_M@gmail.com  364-656-8427  ************5677
4                LHines@verizon.com  713-226-5883  ************5498
…                               …         …             …
119385          Claudia.J@yahoo.com  403-092-5582  ************8647
119386         WAguilar@xfinity.com  238-763-0612  ************4333
119387     Mary_Morales@hotmail.com  395-518-4100  ************1821
119388      MD_Caroline@comcast.net  531-528-1017  ************7860
119389         Ariana_M@xfinity.com  422-804-6403  ************4482

[119390 rows x 36 columns]
```

[3]:
```
df.head()
```

[3]:
```
          hotel  is_canceled  lead_time  arrival_date_year arrival_date_month  \
0  Resort Hotel            0        342               2015               July
1  Resort Hotel            0        737               2015               July
2  Resort Hotel            0          7               2015               July
3  Resort Hotel            0         13               2015               July
4  Resort Hotel            0         14               2015               July

   arrival_date_week_number  arrival_date_day_of_month  \
0                        27                          1
1                        27                          1
2                        27                          1
3                        27                          1
4                        27                          1

   stays_in_weekend_nights  stays_in_week_nights  adults  …  customer_type  \
0                        0                     0       2  …      Transient
1                        0                     0       2  …      Transient
2                        0                     1       1  …      Transient
3                        0                     1       1  …      Transient
4                        0                     2       2  …      Transient

    adr required_car_parking_spaces total_of_special_requests  \
0   0.0                           0                         0
1   0.0                           0                         0
2  75.0                           0                         0
```

```
3  75.0                          0                          0
4  98.0                          0                          1

  reservation_status reservation_status_date           name  \
0          Check-Out              2015-07-01   Ernest Barnes
1          Check-Out              2015-07-01    Andrea Baker
2          Check-Out              2015-07-02  Rebecca Parker
3          Check-Out              2015-07-02    Laura Murray
4          Check-Out              2015-07-03     Linda Hines

                         email  phone-number       credit_card
0  Ernest.Barnes31@outlook.com  669-792-1661  ************4322
1       Andrea_Baker94@aol.com  858-637-6955  ************9157
2   Rebecca_Parker@comcast.net  652-885-2745  ************3734
3            Laura_M@gmail.com  364-656-8427  ************5677
4           LHines@verizon.com  713-226-5883  ************5498

[5 rows x 36 columns]
```

[4]:
```
df.shape
```

[4]:
```
(119390, 36)
```

[5]:
```
df.columns
```

[5]:
```
Index(['hotel', 'is_canceled', 'lead_time', 'arrival_date_year',
       'arrival_date_month', 'arrival_date_week_number',
       'arrival_date_day_of_month', 'stays_in_weekend_nights',
       'stays_in_week_nights', 'adults', 'children', 'babies', 'meal',
       'country', 'market_segment', 'distribution_channel',
       'is_repeated_guest', 'previous_cancellations',
       'previous_bookings_not_canceled', 'reserved_room_type',
       'assigned_room_type', 'booking_changes', 'deposit_type', 'agent',
       'company', 'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date', 'name', 'email',
       'phone-number', 'credit_card'],
      dtype='object')
```

[6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 36 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   hotel                           119390 non-null  object
 1   is_canceled                     119390 non-null  int64
```

```
 2   lead_time                       119390 non-null  int64
 3   arrival_date_year               119390 non-null  int64
 4   arrival_date_month              119390 non-null  object
 5   arrival_date_week_number        119390 non-null  int64
 6   arrival_date_day_of_month       119390 non-null  int64
 7   stays_in_weekend_nights         119390 non-null  int64
 8   stays_in_week_nights            119390 non-null  int64
 9   adults                          119390 non-null  int64
 10  children                        119386 non-null  float64
 11  babies                          119390 non-null  int64
 12  meal                            119390 non-null  object
 13  country                         118902 non-null  object
 14  market_segment                  119390 non-null  object
 15  distribution_channel            119390 non-null  object
 16  is_repeated_guest               119390 non-null  int64
 17  previous_cancellations          119390 non-null  int64
 18  previous_bookings_not_canceled  119390 non-null  int64
 19  reserved_room_type              119390 non-null  object
 20  assigned_room_type              119390 non-null  object
 21  booking_changes                 119390 non-null  int64
 22  deposit_type                    119390 non-null  object
 23  agent                           103050 non-null  float64
 24  company                         6797 non-null    float64
 25  days_in_waiting_list            119390 non-null  int64
 26  customer_type                   119390 non-null  object
 27  adr                             119390 non-null  float64
 28  required_car_parking_spaces     119390 non-null  int64
 29  total_of_special_requests       119390 non-null  int64
 30  reservation_status              119390 non-null  object
 31  reservation_status_date         119390 non-null  object
 32  name                            119390 non-null  object
 33  email                           119390 non-null  object
 34  phone-number                    119390 non-null  object
 35  credit_card                     119390 non-null  object
dtypes: float64(4), int64(16), object(16)
memory usage: 32.8+ MB
```

[7]:
```
#Converting reservation_status _date into datetime format
df['reservation_status_date']=pd.to_datetime(df['reservation_status_date'])
```

[8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 36 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   hotel                           119390 non-null  object
```

```
 1   is_canceled                     119390 non-null  int64
 2   lead_time                       119390 non-null  int64
 3   arrival_date_year               119390 non-null  int64
 4   arrival_date_month              119390 non-null  object
 5   arrival_date_week_number        119390 non-null  int64
 6   arrival_date_day_of_month       119390 non-null  int64
 7   stays_in_weekend_nights         119390 non-null  int64
 8   stays_in_week_nights            119390 non-null  int64
 9   adults                          119390 non-null  int64
 10  children                        119386 non-null  float64
 11  babies                          119390 non-null  int64
 12  meal                            119390 non-null  object
 13  country                         118902 non-null  object
 14  market_segment                  119390 non-null  object
 15  distribution_channel            119390 non-null  object
 16  is_repeated_guest               119390 non-null  int64
 17  previous_cancellations          119390 non-null  int64
 18  previous_bookings_not_canceled  119390 non-null  int64
 19  reserved_room_type              119390 non-null  object
 20  assigned_room_type              119390 non-null  object
 21  booking_changes                 119390 non-null  int64
 22  deposit_type                    119390 non-null  object
 23  agent                           103050 non-null  float64
 24  company                         6797 non-null    float64
 25  days_in_waiting_list            119390 non-null  int64
 26  customer_type                   119390 non-null  object
 27  adr                             119390 non-null  float64
 28  required_car_parking_spaces     119390 non-null  int64
 29  total_of_special_requests       119390 non-null  int64
 30  reservation_status              119390 non-null  object
 31  reservation_status_date         119390 non-null  datetime64[ns]
 32  name                            119390 non-null  object
 33  email                           119390 non-null  object
 34  phone-number                    119390 non-null  object
 35  credit_card                     119390 non-null  object
dtypes: datetime64[ns](1), float64(4), int64(16), object(15)
memory usage: 32.8+ MB
```

[9]:
```
df.describe()
```

[9]:
```
          is_canceled      lead_time  arrival_date_year  \
count  119390.000000  119390.000000      119390.000000
mean        0.370416     104.011416        2016.156554
std         0.482918     106.863097           0.707476
min         0.000000       0.000000        2015.000000
25%         0.000000      18.000000        2016.000000
50%         0.000000      69.000000        2016.000000
```

```
75%         1.000000     160.000000        2017.000000
max         1.000000     737.000000        2017.000000

       arrival_date_week_number  arrival_date_day_of_month  \
count             119390.000000              119390.000000
mean                  27.165173                  15.798241
std                   13.605138                   8.780829
min                    1.000000                   1.000000
25%                   16.000000                   8.000000
50%                   28.000000                  16.000000
75%                   38.000000                  23.000000
max                   53.000000                  31.000000

       stays_in_weekend_nights  stays_in_week_nights         adults  \
count            119390.000000         119390.000000  119390.000000
mean                  0.927599              2.500302       1.856403
std                   0.998613              1.908286       0.579261
min                   0.000000              0.000000       0.000000
25%                   0.000000              1.000000       2.000000
50%                   1.000000              2.000000       2.000000
75%                   2.000000              3.000000       2.000000
max                  19.000000             50.000000      55.000000

            children         babies  is_repeated_guest  \
count  119386.000000  119390.000000      119390.000000
mean        0.103890       0.007949           0.031912
std         0.398561       0.097436           0.175767
min         0.000000       0.000000           0.000000
25%         0.000000       0.000000           0.000000
50%         0.000000       0.000000           0.000000
75%         0.000000       0.000000           0.000000
max        10.000000      10.000000           1.000000

       previous_cancellations  previous_bookings_not_canceled  \
count           119390.000000                   119390.000000
mean                 0.087118                        0.137097
std                  0.844336                        1.497437
min                  0.000000                        0.000000
25%                  0.000000                        0.000000
50%                  0.000000                        0.000000
75%                  0.000000                        0.000000
max                 26.000000                       72.000000

       booking_changes          agent      company  days_in_waiting_list  \
count    119390.000000  103050.000000  6797.000000         119390.000000
mean          0.221124      86.693382   189.266735              2.321149
std           0.652306     110.774548   131.655015             17.594721
```

```
min           0.000000       1.000000     6.000000              0.000000
25%           0.000000       9.000000    62.000000              0.000000
50%           0.000000      14.000000   179.000000              0.000000
75%           0.000000     229.000000   270.000000              0.000000
max          21.000000     535.000000   543.000000            391.000000

                adr  required_car_parking_spaces  total_of_special_requests
count  119390.000000                119390.000000              119390.000000
mean      101.831122                     0.062518                   0.571363
std        50.535790                     0.245291                   0.792798
min        -6.380000                     0.000000                   0.000000
25%        69.290000                     0.000000                   0.000000
50%        94.575000                     0.000000                   0.000000
75%       126.000000                     0.000000                   1.000000
max      5400.000000                     8.000000                   5.000000
```

[10]:
```
#Getting details about only catagorical columns.
df.describe(include='object')
```

[10]:
```
              hotel arrival_date_month    meal country market_segment  \
count        119390              119390  119390  118902         119390
unique            2                  12       5     177              8
top     City Hotel              August      BB     PRT      Online TA
freq          79330               13877   92310   48590          56477

       distribution_channel reserved_room_type assigned_room_type  \
count                119390             119390              119390
unique                    5                 10                  12
top                   TA/TO                  A                   A
freq                  97870              85994               74053

       deposit_type customer_type reservation_status              name  \
count        119390        119390             119390            119390
unique            3             4                  3             81503
top      No Deposit     Transient          Check-Out  Michael Johnson
freq         104641         89613              75166                48

                      email  phone-number       credit_card
count                119390        119390            119390
unique               115889        119390              9000
top     Michael.C@gmail.com  669-792-1661  ************4923
freq                     6             1                28
```

[11]:
```
#Checking the names of catagorical cols.
df.describe(include='object').columns
```

[11]:
```
Index(['hotel', 'arrival_date_month', 'meal', 'country', 'market_segment',
       'distribution_channel', 'reserved_room_type', 'assigned_room_type',
       'deposit_type', 'customer_type', 'reservation_status', 'name', 'email',
       'phone-number', 'credit_card'],
      dtype='object')
```

[12]:
```
# Seeing the data present in the cata cols
for col in df.describe(include='object').columns:
    print(df[col].unique())
    print("-"*50)
```

```
['Resort Hotel' 'City Hotel']
--------------------------------------------------
['July' 'August' 'September' 'October' 'November' 'December' 'January'
 'February' 'March' 'April' 'May' 'June']
--------------------------------------------------
['BB' 'FB' 'HB' 'SC' 'Undefined']
--------------------------------------------------
['PRT' 'GBR' 'USA' 'ESP' 'IRL' 'FRA' nan 'ROU' 'NOR' 'OMN' 'ARG' 'POL'
 'DEU' 'BEL' 'CHE' 'CN' 'GRC' 'ITA' 'NLD' 'DNK' 'RUS' 'SWE' 'AUS' 'EST'
 'CZE' 'BRA' 'FIN' 'MOZ' 'BWA' 'LUX' 'SVN' 'ALB' 'IND' 'CHN' 'MEX' 'MAR'
 'UKR' 'SMR' 'LVA' 'PRI' 'SRB' 'CHL' 'AUT' 'BLR' 'LTU' 'TUR' 'ZAF' 'AGO'
 'ISR' 'CYM' 'ZMB' 'CPV' 'ZWE' 'DZA' 'KOR' 'CRI' 'HUN' 'ARE' 'TUN' 'JAM'
 'HRV' 'HKG' 'IRN' 'GEO' 'AND' 'GIB' 'URY' 'JEY' 'CAF' 'CYP' 'COL' 'GGY'
 'KWT' 'NGA' 'MDV' 'VEN' 'SVK' 'FJI' 'KAZ' 'PAK' 'IDN' 'LBN' 'PHL' 'SEN'
 'SYC' 'AZE' 'BHR' 'NZL' 'THA' 'DOM' 'MKD' 'MYS' 'ARM' 'JPN' 'LKA' 'CUB'
 'CMR' 'BIH' 'MUS' 'COM' 'SUR' 'UGA' 'BGR' 'CIV' 'JOR' 'SYR' 'SGP' 'BDI'
 'SAU' 'VNM' 'PLW' 'QAT' 'EGY' 'PER' 'MLT' 'MWI' 'ECU' 'MDG' 'ISL' 'UZB'
 'NPL' 'BHS' 'MAC' 'TGO' 'TWN' 'DJI' 'STP' 'KNA' 'ETH' 'IRQ' 'HND' 'RWA'
 'KHM' 'MCO' 'BGD' 'IMN' 'TJK' 'NIC' 'BEN' 'VGB' 'TZA' 'GAB' 'GHA' 'TMP'
 'GLP' 'KEN' 'LIE' 'GNB' 'MNE' 'UMI' 'MYT' 'FRO' 'MMR' 'PAN' 'BFA' 'LBY'
 'MLI' 'NAM' 'BOL' 'PRY' 'BRB' 'ABW' 'AIA' 'SLV' 'DMA' 'PYF' 'GUY' 'LCA'
 'ATA' 'GTM' 'ASM' 'MRT' 'NCL' 'KIR' 'SDN' 'ATF' 'SLE' 'LAO']
--------------------------------------------------
['Direct' 'Corporate' 'Online TA' 'Offline TA/TO' 'Complementary' 'Groups'
 'Undefined' 'Aviation']
--------------------------------------------------
['Direct' 'Corporate' 'TA/TO' 'Undefined' 'GDS']
--------------------------------------------------
['C' 'A' 'D' 'E' 'G' 'F' 'H' 'L' 'P' 'B']
--------------------------------------------------
['C' 'A' 'D' 'E' 'G' 'F' 'I' 'B' 'H' 'P' 'L' 'K']
--------------------------------------------------
['No Deposit' 'Refundable' 'Non Refund']
--------------------------------------------------
['Transient' 'Contract' 'Transient-Party' 'Group']
--------------------------------------------------
['Check-Out' 'Canceled' 'No-Show']
```

```
--------------------------------------------------
['Ernest Barnes' 'Andrea Baker' 'Rebecca Parker' … 'Wesley Aguilar'
 'Caroline Conley MD' 'Ariana Michael']
--------------------------------------------------
['Ernest.Barnes31@outlook.com' 'Andrea_Baker94@aol.com'
 'Rebecca_Parker@comcast.net' … 'Mary_Morales@hotmail.com'
 'MD_Caroline@comcast.net' 'Ariana_M@xfinity.com']
--------------------------------------------------
['669-792-1661' '858-637-6955' '652-885-2745' … '395-518-4100'
 '531-528-1017' '422-804-6403']
--------------------------------------------------
['************4322' '************9157' '************3734' …
 '************9170' '************6349' '************7959']
--------------------------------------------------
```

[13]:
```
df.isnull().sum()
```

[13]:
```
hotel                               0
is_canceled                         0
lead_time                           0
arrival_date_year                   0
arrival_date_month                  0
arrival_date_week_number            0
arrival_date_day_of_month           0
stays_in_weekend_nights             0
stays_in_week_nights                0
adults                              0
children                            4
babies                              0
meal                                0
country                           488
market_segment                      0
distribution_channel                0
is_repeated_guest                   0
previous_cancellations              0
previous_bookings_not_canceled      0
reserved_room_type                  0
assigned_room_type                  0
booking_changes                     0
deposit_type                        0
agent                           16340
company                        112593
days_in_waiting_list                0
customer_type                       0
adr                                 0
required_car_parking_spaces         0
total_of_special_requests           0
```

```
reservation_status                  0
reservation_status_date             0
name                                0
email                               0
phone-number                        0
credit_card                         0
dtype: int64
```

[14]:
```
df.drop(['country','agent'], axis=1, inplace = True)
```

[15]:
```
df.drop(['name','email','phone-number','credit_card'], axis=1, inplace=True)
```

[16]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 30 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   hotel                           119390 non-null  object
 1   is_canceled                     119390 non-null  int64
 2   lead_time                       119390 non-null  int64
 3   arrival_date_year               119390 non-null  int64
 4   arrival_date_month              119390 non-null  object
 5   arrival_date_week_number        119390 non-null  int64
 6   arrival_date_day_of_month       119390 non-null  int64
 7   stays_in_weekend_nights         119390 non-null  int64
 8   stays_in_week_nights            119390 non-null  int64
 9   adults                          119390 non-null  int64
 10  children                        119386 non-null  float64
 11  babies                          119390 non-null  int64
 12  meal                            119390 non-null  object
 13  market_segment                  119390 non-null  object
 14  distribution_channel            119390 non-null  object
 15  is_repeated_guest               119390 non-null  int64
 16  previous_cancellations          119390 non-null  int64
 17  previous_bookings_not_canceled  119390 non-null  int64
 18  reserved_room_type              119390 non-null  object
 19  assigned_room_type              119390 non-null  object
 20  booking_changes                 119390 non-null  int64
 21  deposit_type                    119390 non-null  object
 22  company                         6797 non-null    float64
 23  days_in_waiting_list            119390 non-null  int64
 24  customer_type                   119390 non-null  object
 25  adr                             119390 non-null  float64
 26  required_car_parking_spaces     119390 non-null  int64
 27  total_of_special_requests       119390 non-null  int64
 28  reservation_status              119390 non-null  object
```

```
 29  reservation_status_date         119390 non-null  datetime64[ns]
dtypes: datetime64[ns](1), float64(3), int64(16), object(10)
memory usage: 27.3+ MB
```

[17]:
```
df.isnull().sum()
```

[17]:
```
hotel                                0
is_canceled                          0
lead_time                            0
arrival_date_year                    0
arrival_date_month                   0
arrival_date_week_number             0
arrival_date_day_of_month            0
stays_in_weekend_nights              0
stays_in_week_nights                 0
adults                               0
children                             4
babies                               0
meal                                 0
market_segment                       0
distribution_channel                 0
is_repeated_guest                    0
previous_cancellations               0
previous_bookings_not_canceled       0
reserved_room_type                   0
assigned_room_type                   0
booking_changes                      0
deposit_type                         0
company                         112593
days_in_waiting_list                 0
customer_type                        0
adr                                  0
required_car_parking_spaces          0
total_of_special_requests            0
reservation_status                   0
reservation_status_date              0
dtype: int64
```

[18]:
```
df.describe()
```

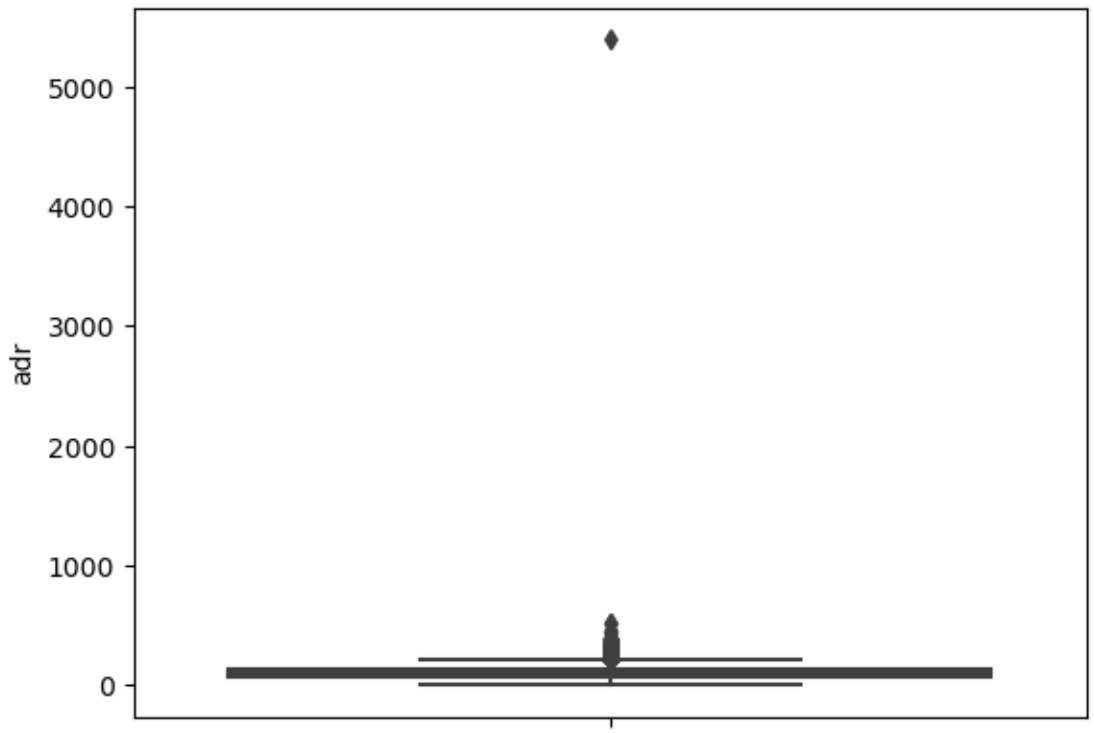
[18]:
```
          is_canceled      lead_time  arrival_date_year  \
count  119390.000000  119390.000000      119390.000000
mean        0.370416     104.011416        2016.156554
std         0.482918     106.863097           0.707476
min         0.000000       0.000000        2015.000000
25%         0.000000      18.000000        2016.000000
50%         0.000000      69.000000        2016.000000
75%         1.000000     160.000000        2017.000000
```

```
max         1.000000     737.000000        2017.000000

       arrival_date_week_number  arrival_date_day_of_month  \
count             119390.000000              119390.000000
mean                  27.165173                  15.798241
std                   13.605138                   8.780829
min                    1.000000                   1.000000
25%                   16.000000                   8.000000
50%                   28.000000                  16.000000
75%                   38.000000                  23.000000
max                   53.000000                  31.000000

       stays_in_weekend_nights  stays_in_week_nights         adults  \
count            119390.000000         119390.000000  119390.000000
mean                  0.927599              2.500302       1.856403
std                   0.998613              1.908286       0.579261
min                   0.000000              0.000000       0.000000
25%                   0.000000              1.000000       2.000000
50%                   1.000000              2.000000       2.000000
75%                   2.000000              3.000000       2.000000
max                  19.000000             50.000000      55.000000

            children         babies  is_repeated_guest  \
count  119386.000000  119390.000000      119390.000000
mean        0.103890       0.007949           0.031912
std         0.398561       0.097436           0.175767
min         0.000000       0.000000           0.000000
25%         0.000000       0.000000           0.000000
50%         0.000000       0.000000           0.000000
75%         0.000000       0.000000           0.000000
max        10.000000      10.000000           1.000000

       previous_cancellations  previous_bookings_not_canceled  \
count           119390.000000                   119390.000000
mean                 0.087118                        0.137097
std                  0.844336                        1.497437
min                  0.000000                        0.000000
25%                  0.000000                        0.000000
50%                  0.000000                        0.000000
75%                  0.000000                        0.000000
max                 26.000000                       72.000000

       booking_changes      company  days_in_waiting_list            adr  \
count    119390.000000  6797.000000         119390.000000  119390.000000
mean          0.221124   189.266735              2.321149     101.831122
std           0.652306   131.655015             17.594721      50.535790
min           0.000000     6.000000              0.000000      -6.380000
```

```
25%            0.000000    62.000000              0.000000      69.290000
50%            0.000000   179.000000              0.000000      94.575000
75%            0.000000   270.000000              0.000000     126.000000
max           21.000000   543.000000            391.000000    5400.000000

       required_car_parking_spaces  total_of_special_requests
count                119390.000000              119390.000000
mean                      0.062518                   0.571363
std                       0.245291                   0.792798
min                       0.000000                   0.000000
25%                       0.000000                   0.000000
50%                       0.000000                   0.000000
75%                       0.000000                   1.000000
max                       8.000000                   5.000000
```

[19]:
```
sns.boxplot(df,y='adr')
```

[19]:
```
<Axes: ylabel='adr'>
```



[20]:
```
df = df[df['adr']<5000]
```

[21]:
```
df.describe()
```

[21]:

	is_canceled	lead_time	arrival_date_year
count	119389.000000	119389.000000	119389.000000
mean	0.370411	104.011994	2016.156555
std	0.482917	106.863358	0.707479
min	0.000000	0.000000	2015.000000
25%	0.000000	18.000000	2016.000000
50%	0.000000	69.000000	2016.000000
75%	1.000000	160.000000	2017.000000
max	1.000000	737.000000	2017.000000

	arrival_date_week_number	arrival_date_day_of_month
count	119389.000000	119389.000000
mean	27.165292	15.798164
std	13.605134	8.780826
min	1.000000	1.000000
25%	16.000000	8.000000
50%	28.000000	16.000000
75%	38.000000	23.000000
max	53.000000	31.000000

	stays_in_weekend_nights	stays_in_week_nights	adults
count	119389.000000	119389.000000	119389.000000
mean	0.927606	2.500314	1.856402
std	0.998614	1.908289	0.579263
min	0.000000	0.000000	0.000000
25%	0.000000	1.000000	2.000000
50%	1.000000	2.000000	2.000000
75%	2.000000	3.000000	2.000000
max	19.000000	50.000000	55.000000

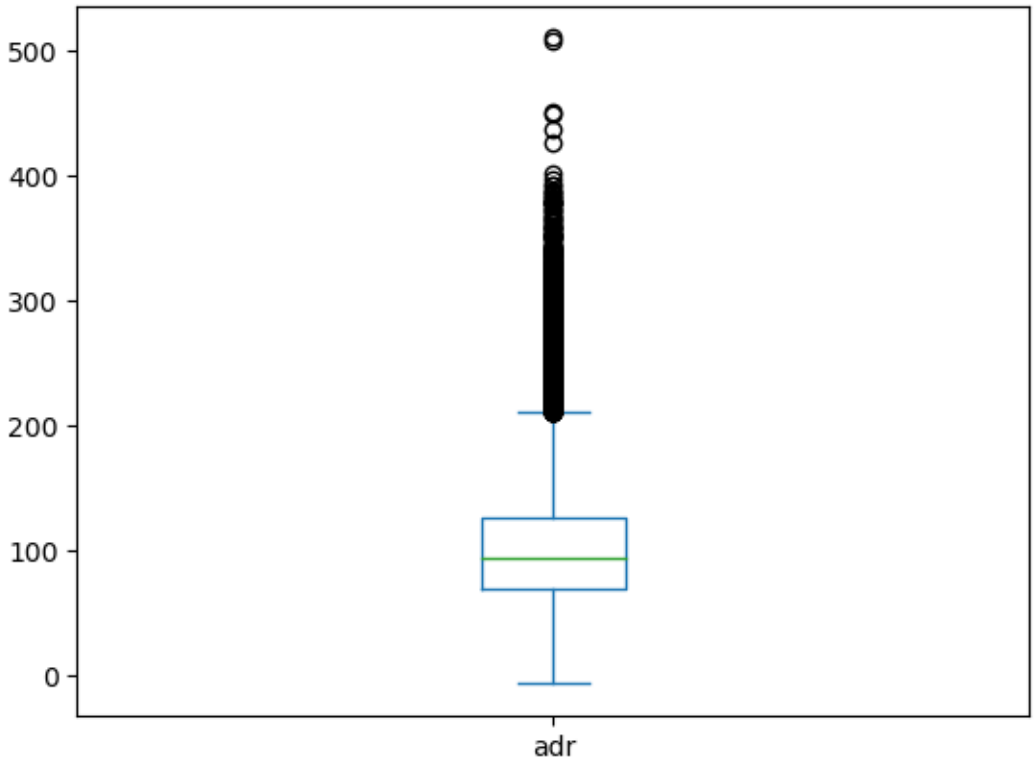
	children	babies	is_repeated_guest
count	119385.000000	119389.000000	119389.000000
mean	0.103891	0.007949	0.031912
std	0.398563	0.097437	0.175768
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000
max	10.000000	10.000000	1.000000

	previous_cancellations	previous_bookings_not_canceled
count	119389.000000	119389.000000
mean	0.087119	0.137098
std	0.844340	1.497443
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000

```
75%                 0.000000                      0.000000
max                26.000000                     72.000000

       booking_changes      company  days_in_waiting_list            adr  \
count    119389.000000  6797.000000         119389.000000  119389.000000
mean          0.221118   189.266735              2.321169     101.786744
std           0.652304   131.655015             17.594793      48.153554
min           0.000000     6.000000              0.000000      -6.380000
25%           0.000000    62.000000              0.000000      69.290000
50%           0.000000   179.000000              0.000000      94.560000
75%           0.000000   270.000000              0.000000     126.000000
max          21.000000   543.000000            391.000000     510.000000

       required_car_parking_spaces  total_of_special_requests
count                119389.000000              119389.000000
mean                      0.062518                   0.571368
std                       0.245292                   0.792800
min                       0.000000                   0.000000
25%                       0.000000                   0.000000
50%                       0.000000                   0.000000
75%                       0.000000                   1.000000
max                       8.000000                   5.000000
```

[22]:
```
df['adr'].plot(kind='box')
```

[22]:
```
<Axes: >
```



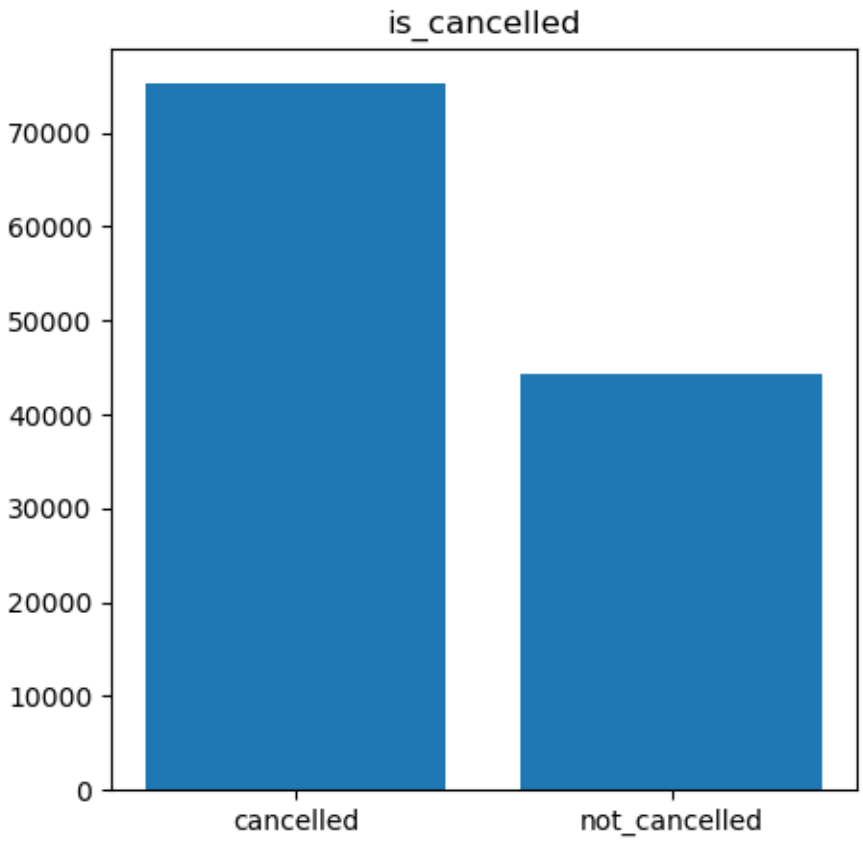
1 Data Analysis and Visualization

[23]:
```
cancelled_perc = df['is_canceled'].value_counts(normalize=True)
cancelled_perc
```

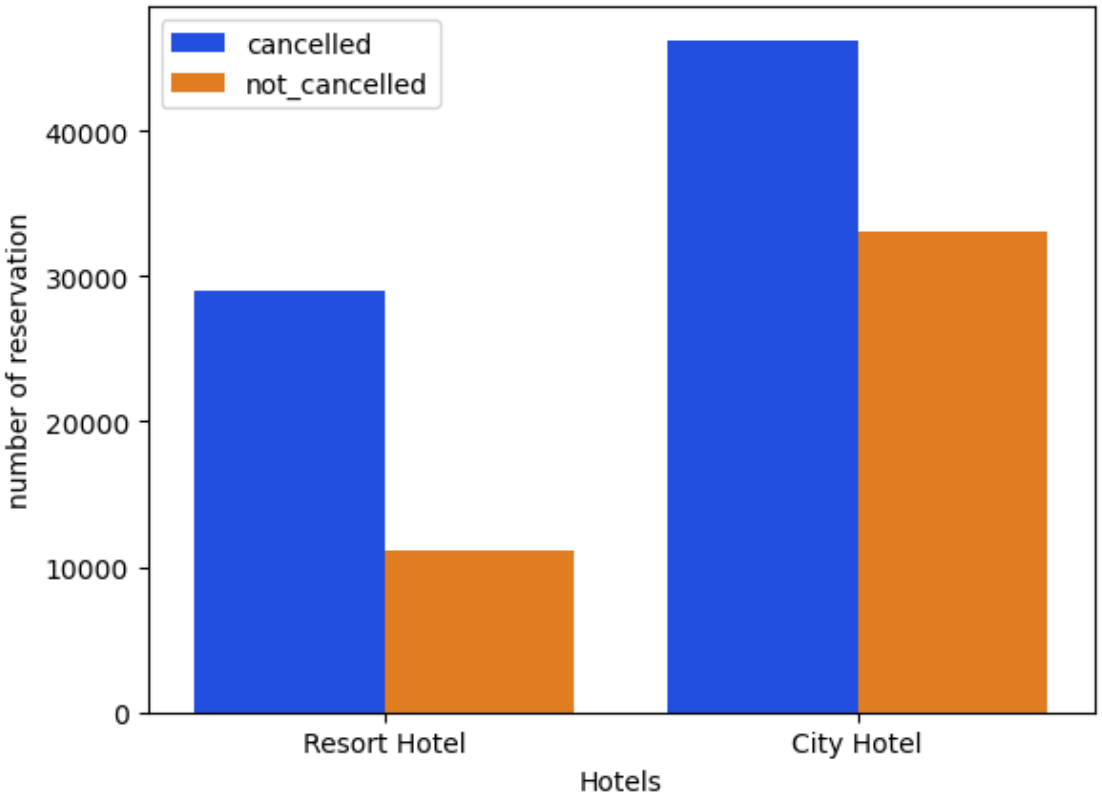
[23]:
```
0    0.629589
1    0.370411
Name: is_canceled, dtype: float64
```

[24]:
```
plt.figure(figsize=(5,5))
plt.title("is_cancelled")
plt.bar(["cancelled","not_cancelled"],df['is_canceled'].value_counts())
```

[24]:
```
<BarContainer object of 2 artists>
```



[25]:
```
#Now we will check in which hotel the cancellation.
cancellation_stats = sns.countplot(data=df, x='hotel', hue='is_canceled',␣
↪palette='bright')
plt.legend(['cancelled',"not_cancelled"])
plt.xlabel('Hotels')
plt.ylabel("number of reservation")
plt.show()
```



[26]:
```
#Now we are checking in resort hotel what is the % of conccellation
resort_hotel = df[df['hotel']=='Resort Hotel']
city_hotel = df[df['hotel']=='City Hotel']
```

[27]:
```
city_hotel['is_canceled'].value_counts(normalize=True)
```

[27]:
```
0    0.582738
1    0.417262
Name: is_canceled, dtype: float64
```

[28]:
```
resort_hotel['is_canceled'].value_counts(normalize=True)
```

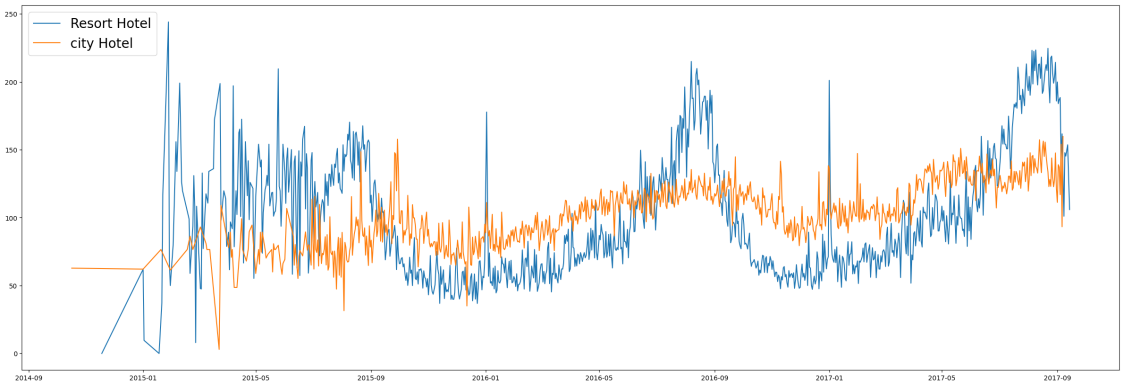
[28]:
```
0    0.722366
1    0.277634
Name: is_canceled, dtype: float64
```

[29]:
```
resort_hotel= resort_hotel.groupby('reservation_status_date')[['adr']].mean()
city_hotel = city_hotel.groupby('reservation_status_date')[['adr']].mean()
```

[30]:
```
plt.figure(figsize=(30,10))
plt.plot(resort_hotel.index,resort_hotel['adr'], label = 'Resort Hotel')
```

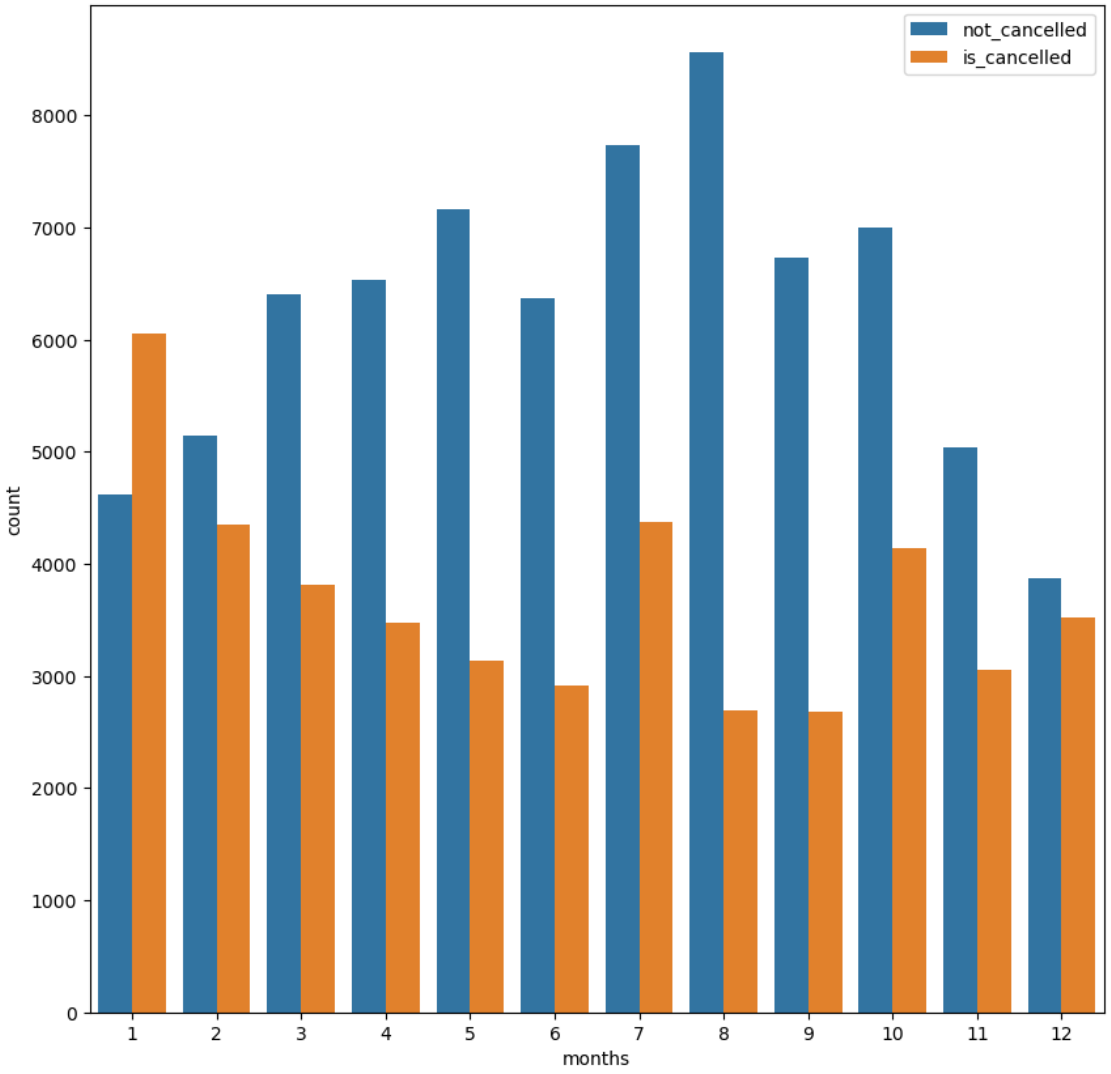
```
plt.plot(city_hotel.index,city_hotel['adr'], label = 'city Hotel')
plt.legend(fontsize=20)
```

[30]: <matplotlib.legend.Legend at 0x1f59328dc00>

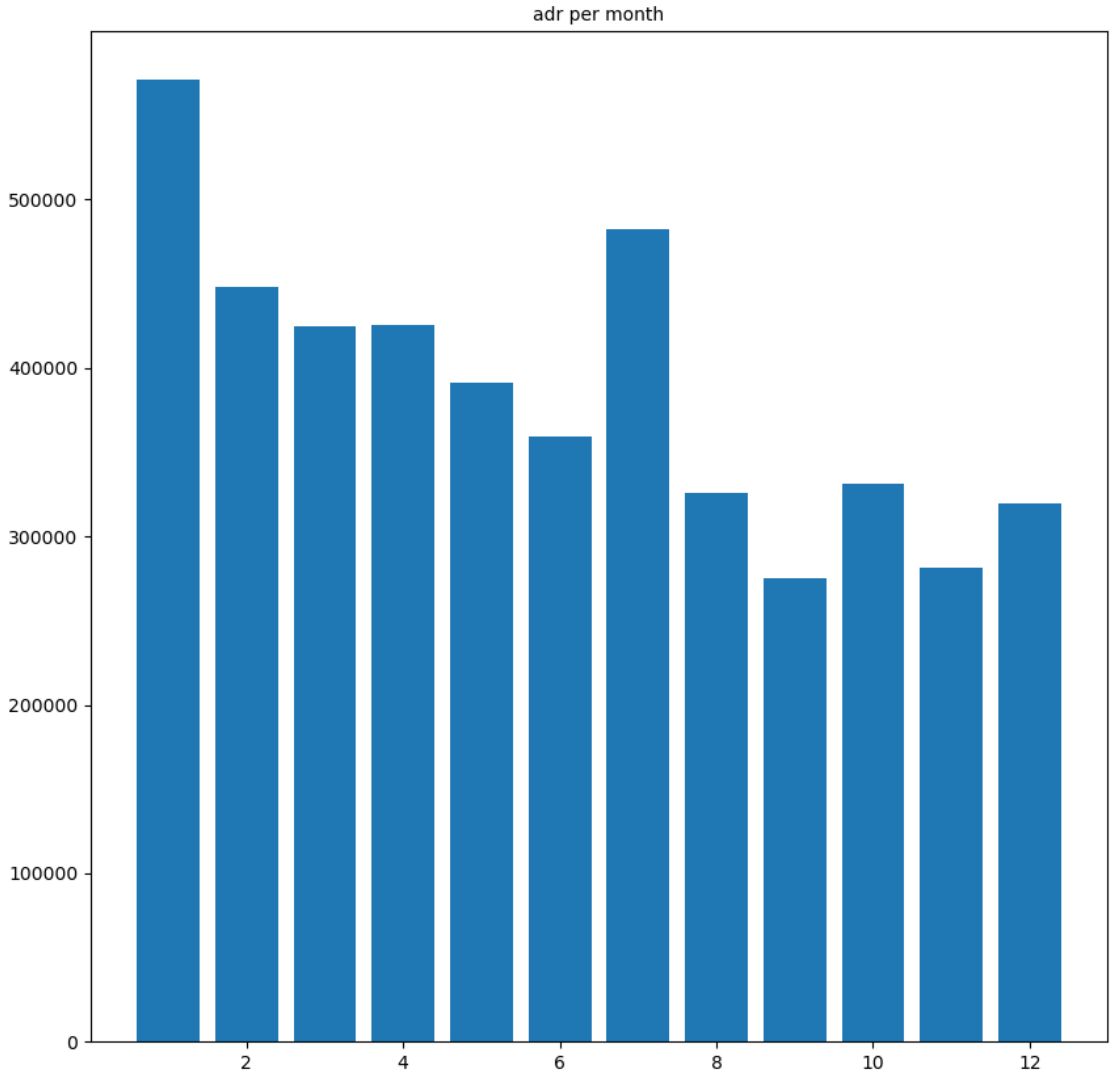


[31]:
```
#In whcih month the reservation happens the most.
plt.figure(figsize=(10,10))
df['months']=df['reservation_status_date'].dt.month
month_status = sns.countplot(data=df, x='months', hue='is_canceled')
plt.legend(['not_cancelled','is_cancelled'])
```

[31]: <matplotlib.legend.Legend at 0x1f593303190>



[32]:
```
#Comparing the price vs per month cancellation rate
plt.figure(figsize=(10,10))
plt.title("adr per month", fontsize=10)
plt.bar('months','adr', data=df[df['is_canceled']==1].
  ↪groupby('months')[['adr']].sum().reset_index())
plt.show()
```

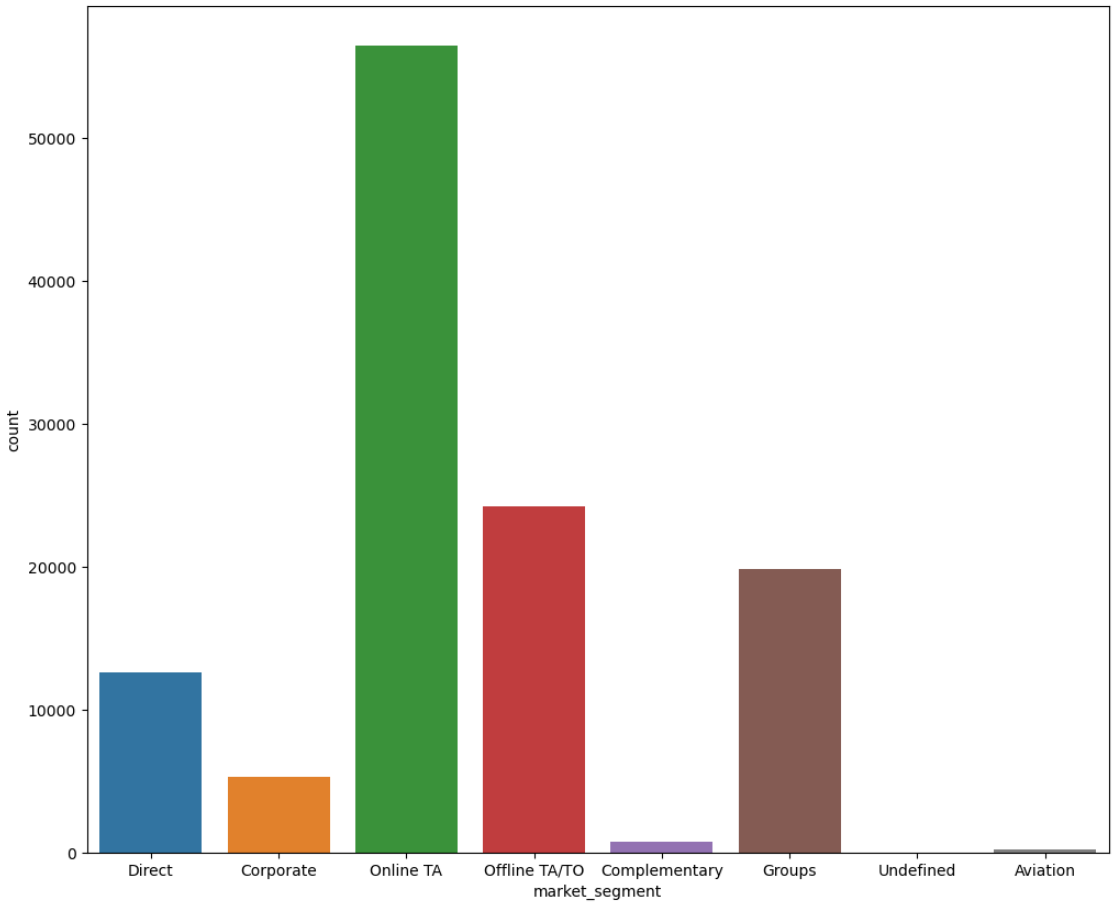


[33]:
```
#Checking through which source the people are coming to the hotel.
df['market_segment'].value_counts(normalize=True)
```

[33]:
```
Online TA         0.473050
Offline TA/TO     0.202850
Groups            0.165937
Direct            0.105588
Corporate         0.044351
Complementary     0.006223
Aviation          0.001985
Undefined         0.000017
Name: market_segment, dtype: float64
```

[34]:
```
plt.figure(figsize=(12,10))
sns.countplot(x='market_segment', data=df)
```

[34]: <Axes: xlabel='market_segment', ylabel='count'>



[ ]: